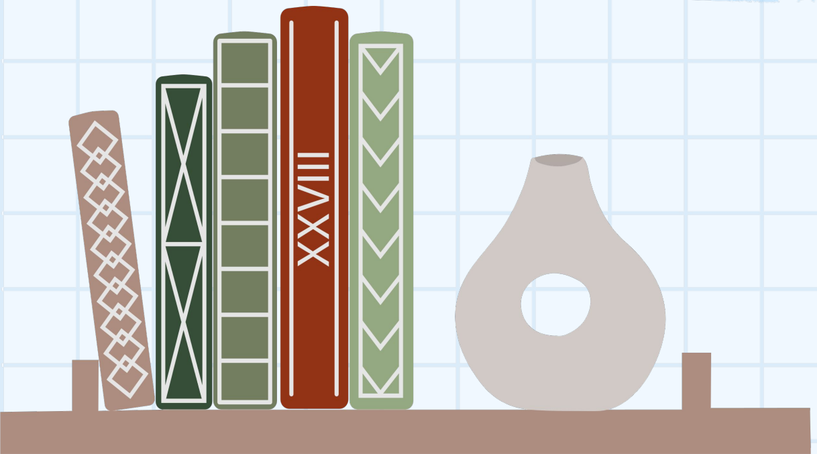
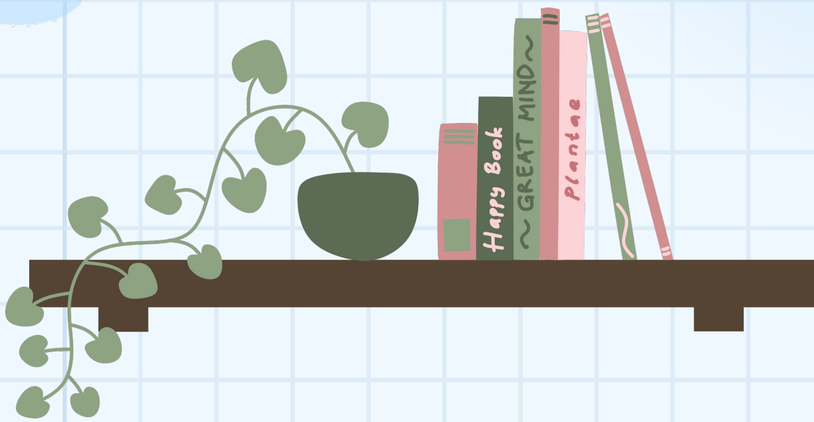
Powered by
Futurense

BS./BSC.IN

Applied AI and Data Science

Basics of Data Analytics





Powered by Futurense

Let's dive into and learn:

1. Data Privacy
2. Data Security
3. Data Ethics

Data Privacy

- Data privacy refers to the principles that govern the collection, use, storage, and sharing of personal data.
- Is important for
  - preventing identity theft
  - building consumer trust
  - building brand image
  - complying with legal requirements
  - promoting ethical practices

Data Privacy

- Data privacy applies to data that can uniquely identify an individual.
- This type of data requires confidential handling.
- Can be of the following types
    - Personally Identifiable Information (PII)
    - Personal Information (PI)
    - Sensitive Personal Information

Personally Identifiable

- This refers to data that can uniquely identify an individual’s identity.
- Example: full name,

  phone number,

  email address,

  Aadhar or PAN number,

  date of birth.

Personal Information

- This includes all PII and additional information that can be linked directly or indirectly to an individual or household,
- Examples : IP addresses,

  geo-locations,

  videos, and

  criminal and case records.

Sensitive Information

- This type of data, when combined with other information, can be linked to an individual and potentially cause harm
- Examples : genetic information

  political beliefs

  religious beliefs

Data Anonymization

- Data anonymization refers to the process of modifying a dataset in such a way that it becomes impossible or very difficult to identify individuals based on the available data.
- Essentially, this means removing or transforming personally identifiable information (PII) from datasets, but still retaining the utility of the data for analysis.

Data Anonymization – Removing

Powered by

Futurense

- Removing or transforming personally identifiable information (PII) from datasets, but still retaining the utility of the data for analysis.
- E.g. Removing names and replacing with unique IDs
- Minimizes the risk of data leakages and re-identification, allowing us to share and analyze data safely without compromising individual privacy.

Data Anonymization

- Removing Personally Identifiable Information

Name	Age	Salary	Location
Priya	29	55000	21 MG road, Patna, 800011
Rahul	32	67000	32/C East Street, Mohali, 140301
Devesh	28	52000	73/A Temple Road, Bhubaneshwar, 751003
Bhavna	31	64000	19 Main Street, Thrissur, 620680

ID	Age	Salary	Location
GH89765	29	55000	21 MG road, Patna, 800011
RF23876	32	67000	32/C East Street, Mohali, 140301
UT75097	28	52000	73/A Temple Road, Bhubaneshwar, 751003
FC37694	31	64000	19 Main Street, Thrissur, 620680

Data Anonymization -

- Removing granularity
- Instead of removing data, generalization transforms it into a broader, less identifiable form
- Generalization reduces the granularity of data to prevent identification.
- Data remains useful for analysis but lowers the risk of re-identification.
- E.g. Replacing detailed addresses with broader location, Replacing exact birth date with the age

Data Anonymization

Powered by

Futurense

- Generalization

Name	Age	Salary	Location
Priya	29	55000	21 MG road, Patna, 800011
Rahul	32	67000	32/C East Street, Mohali, 140301
Devesh	28	52000	73/A Temple Road, Bhubaneshwar, 751003
Bhavna	31	64000	19 Main Street, Thrissur, 620680

Name	Age	Salary	Location
Priya	29	55000	Patna
Rahul	32	67000	Mohali
Devesh	28	52000	Bhubaneshwar
Bhavna	31	64000	Thrissur

Data Anonymization –

- When exact value of the data point is not necessary but the overall distribution
- Process of modifying the original data in a controlled manner to protect privacy
- Include various techniques like randomization, scaling, or swapping values.
- Data perturbation aims to obscure data while retaining its usefulness for analysis.
- E.g. Noise addition : introducing random or systematic changes, often called “noise”, to the data

Data Anonymization

- Data Perturbation – Adding Noise

Name	Age	Salary	Location
Priya	29	55000	21 MG road, Patna, 800011
Rahul	32	67000	32/C East Street, Mohali, 140301
Devesh	28	52000	73/A Temple Road, Bhubaneshwar, 751003
Bhavna	31	64000	19 Main Street, Thrissur, 620680

Name	Age	Salary	Location
Priya	29	55248	21 MG road, Patna, 800011
Rahul	32	66930	32/C East Street, Mohali, 140301
Devesh	28	52560	73/A Temple Road, Bhubaneshwar, 751003
Bhavna	31	63970	19 Main Street, Thrissur, 620680

Data Masking

- Data Masking is another process of maintaining data privacy
- Involves altering or obfuscating the original data while maintaining its format and structure
- Often done for unique Id type variables e.g. Aadhar Number

Data Masking

Name	LAM number	Age	Salary
Priya	234-765-8753	29	55000
Rahul	123-872-2396	32	67000
Devesh	714-964-3428	28	52000
Bhavna	561-983-5627	31	64000

Name	LAM number	Age	Salary
Priya	XXX-XXX-XX53	29	55000
Rahul	XXX-XXX-XX96	32	67000
Devesh	XXX-XXX-XX28	28	52000
Bhavna	XXX-XXX-XX27	31	64000

Data Anonymization vs Data

- Data anonymization involves making data completely unidentifiable by removing or altering personal identifiers
- Data masking involves obscuring sensitive data by replacing it with padding or fictitious values while maintaining the original data format
- Data anonymization is usually irreversible
- Data masking usually reversible

Data Anonymization Tools

- ARX
    - Open-source data anonymization tool, ideal for large organizations with large datasets
- IBM Guardian
    - Designed to protect sensitive data across hybrid, multi-cloud environments.
- Google Tensor Flow Privacy
    - Appropriate for companies or individuals developing the models themselves.

Data Privacy Regulations

- Various laws and regulations govern data privacy
  - General Data Protection Regulation (GDPR) in Europe
  - California Consumer Privacy Act (CCPA) in the US
  - Digital Personal Data Protection (DPDP) Act in India

Data Privacy vs Data Security

Powered by

Futurense

- Data Privacy and Data Security are related but distinct areas
- **Data Privacy** deals with the rights of individuals who own the data
- **Data Security** deals with protecting data from unauthorized access and misuse

Data Privacy vs Data Security

Powered by Futurense

- Data security supports data privacy by ensuring that only authorized individuals can access personal data for legitimate purposes.

- Conversely, data privacy supports data security by defining who those authorized individuals are and what constitutes legitimate purposes for accessing the data

Ensuring Data Security

- Collect only necessary information
    - reduce security risks and ensure privacy by only collecting necessary data critical to your business
- Limit Access to Data
    - determine key team members who need access to your data management systems and give necessary permissions to only those members

Ensuring Data Security

- Create an Incident Response Plan
  - be prepared for a hack or security breach in case it happens
  - create an incident response plan
  - outline the exact steps to take to mitigate any incidents

Effects of AI in Data Privacy

- Similar to other areas of data analytics, AI and ML brings both opportunities and challenges
- AI and ML can enhance data protection by automating threat detection, identifying patterns in large datasets, and improving anomaly detection
- AI models itself raise privacy concerns require vast amounts of data for training, raising questions about consent and data usage.

Data Ethics

- Data Ethics refers to the principles behind how organizations gather, protect, and use data.
- Focuses on the moral obligations that entities have (or should have) when collecting and disseminating information about us.
- Regardless of your technical expertise, very important to handle data ethically!

Principles of Data Ethics

- **Transparency:** Clear communication about data collection, storage, and sharing practices.
- **Accountability:** Organizations taking responsibility for the data they collect, including protecting it from breaches and misuse.
- **Individual Agency**: Individuals having control over their personal data, including the ability to access, correct, or delete their information.
- **Data Privacy**: Personal data of individuals will be protected from unauthorized exposure.

Recap

Powered by Futurense

Concepts of Data Privacy and Data Ethics

Some common ways of maintaining Data

Privacy

Powered by Futurense

Thank you